



Service capacity pooling in M/G/1 service systems

Jing Peng, Oualid Jouini, Yves Dallery, Zied Jemai

► To cite this version:

Jing Peng, Oualid Jouini, Yves Dallery, Zied Jemai. Service capacity pooling in M/G/1 service systems. International Conference on Industrial Engineering and Systems Management, Oct 2015, Seville, Spain. hal-01266002

HAL Id: hal-01266002

<https://hal.science/hal-01266002>

Submitted on 2 Feb 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Service Capacity Pooling in M/G/1 Service Systems

(presented at the 6th IESM Conference, October 2015, Seville, Spain) © I⁴e² 2015

Jing Peng, Oualid Jouini, Yves Dallery
Laboratoire Genie Industriel, CentraleSupélec, Université Paris-Saclay
Grand Voie des Vignes, 92290, Châtenay-Malabry, France
Email: jing.peng@ecp.fr, oualid.jouini@ecp.fr, yves.dallery@ecp.fr

Zied Jemai
OASIS -ENIT,
University of Tunis El Manar
BP 37, Le Belvedere 1002, Tunisia
Email: zied.jemai@ecp.fr

Abstract—We study the cost-sharing problem among independent service providers in a service capacity pooling system. The effective improvement of such pooling system can be achieved by reducing the resource idleness in case of congestion. In this paper, we model both the service provider and the cooperative coalition as a single server queue. We attempt to answer the following questions: (i) which coalition strategy should be used; and (ii) which allocation rule should be selected in order to maintain the stability of the coalition? In particular, we consider the service pooling with a fixed service capacity for M/G/1 service systems. The benefit of the pooling system is due to the shortened waiting queue in the overall system. We develop the corresponding cooperative game with transferable utility, and analyze the core allocations. Although it is difficult to express a core allocation explicitly for the game, we prove the non-emptiness of the core. We give a reasonable expression of Equal Profit Method to distribute the cost for our game, and investigate a number of cost allocation rules under three typical situations to evaluate the gain of the service pooling strategy for each service provider. The numerical results show that the cost allocation rule proposed gives a reasonable cost-sharing result considering the contribution of each participant.

Key words: service pooling; queueing systems; cooperative game theory; core allocations; general service times

I. INTRODUCTION

Services play an important role in the world economy. Based on the report from the office of United States Trade Representative, four out of five jobs in US are proposed by service industry [1]. In France, service presents the biggest sector of the economy and accounts for 79% of total GDP according to statistical data from the World Bank. In the context of economic globalization, competition and cooperation in service industries are more and more popular in nowadays: price competition among fast food restaurant chains, operation combination of telecommunication companies, collaborative after-sell and maintenance services in electronic manufacturing industry etc. In this paper, we focus on the collaboration benefit between several independent service providers in terms of capacity sharing. In order to make service systems more efficient, there are several basic cooperative methods: queueing cooperation, e.g., scheduling among simultaneous arrival agencies or rerouting among different servers [2], [3]; service pooling, e.g., service rate pooling or staffing allocation [4]; cross-training [5]; collaboration with third-party service providers, e.g., service outsourcing [6], etc. In some cases, different methods can be combined to form a more profitable collaborative structure [7]. Under certain conditions, service

providers could earn more revenue or reduce expenses by using the collaboration methods mentioned above.

All the previous methods can be classified broadly into three typical cooperation forms: vertical form, i.e., collaboration between customers and servers; horizontal form, i.e., collaboration among homogeneous servers; and external form, e.g., collaborative outsourcing. We focus in this study on horizontal service pooling among independent service providers. The obvious gain is the mitigated congestion in the cooperative system, owing to the disappearance of servers' idleness with waiting customers in the queues. Altogether, the coalition service set is more efficient than individual ones. The pooling advantage for the whole system is apparent, but the collaborative gain of the entire alliance cannot be the incentive for each individual service provider to join the coalition. It is therefore important to address the following questions: which independent service providers should collaborate together; and how to share the cost of the pooling system among the participants to keep each individual staying in the coalition?

In this paper, we consider a set of independent single-server service providers, each of which faces its own incoming stream of customers/demand. Customer inter-arrival and service times are assumed to be random and distributed independently. We suppose that an individual incoming stream is strictly unrelated to those of other providers. This means that there is no competitive relationship among service providers in the set. The providers in the coalition could then join a profitable coalition by operating their service capacities in common. Alternatively, each provider could make his own decision independently to either join any coalition or not, based on his individual benefit. Once the coalition is formed, the most important problem in the entire coalition for every independent provider returns to a cost-sharing problem under collaboration.

Cooperative game theory provides interesting concepts to find profitable coalition structures and solves the cost sharing problem among cooperative players. We assume that the total cost is a transferable utility, e.g., money in the general case. The corresponding cooperative game with transferable utility (TU-game) is defined among the set of independent service providers, and has a characteristic function defined by the total operating costs associated with each coalition for the cases of fixed system capacity. In order to make a profitable service pooling strategy, we prove that the system total cost could be minimized in the grand coalition. Therefore, as mentioned above, we focus on the cost-sharing problem to guarantee the stability of the grand coalition.

There exists a large literature focusing on the cooperative

behavior among independent participants using cooperative game theory: [8] is the earliest research that deals with the cost allocation problem for operating-theatre sharing in medical service. The authors separate the operating-theatre costs as variable and fixed cost, and focus on the Shapley value of two sub-games. In [9], the authors address an extension of [8], considering preemptive priority for the customers from different individual servers to provide a more profitable pooling system. Our work is inspired by two recent papers [10], [11], which study the service capacity pooling problem using cooperative game theory. Two similar general M/M/1 service pooling systems have been considered in these two papers. [10] uses the optimal service rate that minimizes the system operating cost, in addition, an incomplete information problem has been treated here. In [11], the authors choose a variable service rate, which varies with customer incoming rate. Based on a system performance defined in terms of the mean number of waiting customers in the system, authors give the expressions of all cost allocation rules. With the work of [12], [13], the service cost allocation problem in [10] and [11] are extended to multi-server settings with Erlang-B and Erlang-C queueing models.

Using an M/M/1 queueing modeling, several existing studies address the pooling, scheduling, staff allocation, or outsourcing problems in service systems [2]–[4], [6]–[11]. In this work, we extend the service pooling problem to the more general case of M/G/1 systems. Although its prevalence in practice, no studies of service pooling for M/G/1 systems exist in the literature. Considering the efficiency and quality driven in service systems, our system total cost consists of the service capacity cost and the system congestion cost. We prove that the service pooling game with a fixed service capacity sharing always has a stable cost-sharing allocation for the grand coalition. We attempt to consider the contribution of each service provider to coalitions with the Equal Profit Method proposed by [14]. With numerical experiments, we compare the allocation method proposed using two different contribution weights with the Shapley-value, the Tau-value, the nucleolus and the Equal Profit Method. By a reasonable selection of the contribution weights for each individual participant, this rule gives an attracted pooling gain for the participant with a high contribution to the coalition, especially in the quality driven condition.

The rest of the paper is structured as follows: Section 2 presents both the individual and collaborative modeling of service systems. In Section 3, the service pooling problem with a fixed service capacity is defined and analyzed as a TU-game. Then, we propose a reasonable expression of Equal Profit Method, and investigate the allocation method proposed with a number of traditional cost allocation rules for our game under three typical situations in Section 4. Finally, conclusions and future work are presented.

II. SERVICE POOLING MODELING

We consider a set of n service providers, $N = \{1, \dots, n\}$. Each service provider $i \in N$ is modeled as a single server queue handling a single class of customers. We assume that the waiting space is large enough, no customer would abandon after entering the system, and there is no fail in the service processing, i.e., no retrial is considered here. The incoming stream of customers/demand to the service provider i follows

a Poisson process, and the customers are served in the order of their arrivals, i.e., under the first come, first served (FCFS) discipline of service. The service times for a given service provider are assumed to be identically distributed and allowed to follow a general distribution. Following the above assumptions, the individual service process is modeled as an M/G/1 queueing system. For the total operating cost of an individual server, we consider a traditional economic framework with two types of costs [15]. The first type is a linear capacity cost per unit time, which is proportional to the system service capacity. This captures the equipment's depreciation or maintenance fee, employee's salary, etc. We also assume a congestion cost incurred for each unit time the customers spend in the queue. For each service provider i , we then define the following parameters:

- λ_i : Mean arrival rate of customers for the provider i ;
- X_i : A random variable describing the service time at the server i , with mean $1/\mu_i$ and coefficient of variation c_i ;
- $\rho_i = \lambda_i/\mu_i$: Server utilization for the provider i , with $\mu_i > \lambda_i$;
- $W_{q,i}$: Customer expected waiting time in the queue for the provider i ;
- h_i : Service capacity cost parameter per unit time for the provider i ;
- d_i : Congestion cost parameter per unit time per customer waiting in the queue at the server i .

For any service provider i , we denote the capacity, the congestion and the total costs per unit time by H_i , D_i and C_i respectively. Using the Pollaczek-Khinchin formula in [16], we could write

$$\begin{aligned} C_i &= H_i + D_i = h_i\mu_i + d_i\lambda_i W_{q,i} \\ &= h_i\mu_i + d_i\lambda_i \frac{\lambda_i \mathbf{E}(X_i^2)}{2(1-\rho_i)}. \end{aligned} \quad (1)$$

The service capacity pooling consists of two typical methods with demand pooling. In the first one, the service providers form a common facility with parallel-servers and one single queue. All incoming customers are served in the multi-server queueing system. In the second one, the service providers share their service capacities together and run as a "super-server", i.e., a single server with a high service capacity. We consider there the second configuration. We assume that s independent service providers of any subset $\emptyset \subset S = \{1, \dots, s\} \subseteq N$ would decide to share their capacities as a "super-server" for reducing the collective cost. We suppose that the component arrival processes, with rate λ_i for $i \in S$, are independently distributed. Therefore, the combined arrival process in the pooling system, which has a rate $\lambda_S = \sum_{i \in S} \lambda_i$, also follows a Poisson process. We suppose that the pooling system provides same services to all customers, and the service times for the pooling server follow a general distribution. We denote the mean service capacity by μ_S for the pooling system S . Based on these assumptions, the s providers act as an M/G/1 "super-server" (Figure 1). By using the same technology and paying

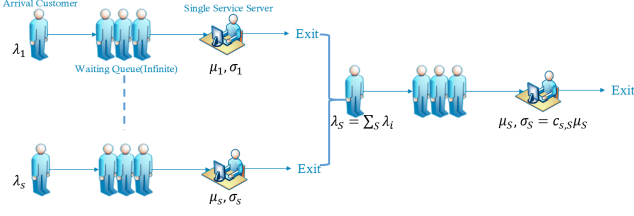


Fig. 1. Centralized Coalition of a Set of Service Providers

at the same salary level, we assume here the service capacity cost parameters to be the same $h_i = h$, for all providers in the set N . Because all the customers have equal importance for the pooling system S , it is reasonable to suppose that the system congestion cost parameters are equal, $d_i = d$, for any provider in the set N . In particular, we assume that the coefficient of variation of service times for all providers in the set N are the equivalent $c_{S,i} = c_S = c_N$. The total cost of the pooling system S is the sum of service capacity cost and system congestion cost as individual providers. Thus, for any pooling system $S \subset N$, we define the parameters as follows:

- λ_S : Mean arrival rate of customers for the pooling system S ;
- X_S : A random variable, service time at the pooling server S , with the mean $1/\mu_S$ and the coefficient of variation c_S ;
- $\rho_S = \lambda_S/\mu_S$: Server utilization for the pooling system S , with $\mu_S > \lambda_S$;
- $W_{q,S}$: Customer expected waiting time in the queue for the pooling system S ;
- h : Service capacity cost parameter per unit time within the subset S ;
- d : Congestion cost parameter per unit time per customer waiting in the queue within the subset S .

For a pooling system S , we denote the capacity, the congestion and the total costs per unit time by H_S , D_S and C_S , respectively. The expected total cost per unit time is given by

$$\begin{aligned} C_S &= H_S + D_S = h\mu_S + d\lambda_S W_{q,S} \\ &= h\mu_S + d\lambda_S \frac{\lambda_S \mathbf{E}(X_S^2)}{2(1 - \rho_S)}. \end{aligned} \quad (2)$$

With all definitions above, we propose both the individual service model and the pooling service model in M/G/1 queueing systems. Our objective is to verify the benefit and the stability of service capacity pooling. It consists of two steps: to confirm that the service capacity pooling brings benefit to the whole set of service providers; to verify that each individual service provider could get better interest in the coalition structure with its distributed cost from the coalition cost. Cooperative game theory provides a formal method to formulate, structure and analyze the cooperative behavior of independent individuals. In the two following sections, we will build the corresponding cooperative game under the capacity pooling strategy, in order to find the most profitable pooling coalition structure and allocate the total cost of pooling server to each independent service provider.

III. SERVICE POOLING GAME FOR M/G/1 SERVICE SYSTEMS

In this section, we consider that the service capacity of each individual service provider is fixed. This is a reasonable assumption in several situations, where the changing of basic equipment or physical location is too expensive or almost impossible. We suppose that the pooling capacity μ_S of any subset S is the sum of service capacities of all members $i \in S$. Let us denote by $C_{fix}(\cdot)$, the total cost function of a subset $S \subset N$ under the fixed service capacity pooling. With the predetermined system parameters (h, d, c_N) , we have

$$\begin{aligned} C_{fix}(\emptyset) &= 0, C_{fix}(i) = C_i(\lambda_i, \mu_i), \\ C_{fix}(S) &= C_S(\lambda_S, \mu_S = \sum_{j \in S} \mu_j). \end{aligned} \quad (3)$$

Consider a finite set of independent service providers $N = \{1, \dots, n\}$ (agents), which is known as the grand coalition in cooperative game. Let S be any subset of N , and it is called coalition of agents. In every coalition, the agents could reduce their total costs by sharing their capacities together. We assume that each service provider could only join one coalition, and the pooling cost could be redistributed among the providers with no limitation in the coalition. Thus, a cooperative game with transferable utility (TU-game) (N, C_{fix}) is completely specified by its characteristic function $C_{fix} : 2^N \rightarrow \mathbf{R}$ assigning to any non-empty coalition S .

To simplify the analysis of the game (N, C_{fix}) , we define $f = c_N^2$ and rewrite equations (1) and (2) as follows.

$$\begin{aligned} C_i &= h_i \mu_i + d_i \lambda_i \frac{\lambda_i \mathbf{E}(X_i^2)}{2(1 - \rho_i)} \\ &= h\mu_i + \frac{d(1+f)}{2\rho_i^{-1}(\rho_i^{-1} - 1)}, \text{ for any } i \in N; \end{aligned} \quad (4a)$$

$$\begin{aligned} C_S &= h\mu_S + d\lambda_S \frac{\lambda_S \mathbf{E}(X_S^2)}{2(1 - \rho_S)} \\ &= h\mu_S + \frac{d(1+f)}{2\rho_S^{-1}(\rho_S^{-1} - 1)}, \text{ for any } \emptyset \subseteq S \subseteq N. \end{aligned} \quad (4b)$$

Note that the capacity holding cost and the system congestion cost, both are single-variable functions in either the service capacity, or the server utilization. We denote the expected queue length of coalition S by $L_{q,S}$. Proposition 1 states the effectiveness improvement of service quality in terms of $L_{q,S}$ in the pooling system.

Proposition 1. The expected queue length is strictly subadditive in the pooling system with a fixed service capacity.

Proof. For any $\emptyset \subseteq S, T \subseteq N$ with $S \cap T = \emptyset$, we suppose that $\rho_S \leq \rho_T$. Then, the utilization of the pooling server $S \cup T$ has the property: $\rho_S \leq \rho_{S \cup T} \leq \rho_T$. And the queue length $L_{q,S} = (1+f)/[2\rho_S^{-1}(\rho_S^{-1} - 1)]$ is an increasing function in $0 < \rho_S < 1$. Thus, $L_{q,S} \leq L_{q,S \cup T} \leq L_{q,T}$. So $L_{q,S \cup T} < L_{q,S} + L_{q,T}$. The subadditivity of expected queue length has been proved. \square

From Proposition 1, we find that the average number of waiting customers is always reduced in the pooling system. The proof above also shows that $L_{q,S}$ is not monotone, the

joint queue length may be increased with new number joining, e.g., when provider i joins coalition S with $\rho_S < \rho_i$. This property is similar to that in the M/M/1 service pooling game proposed by Anily in [11]. We now present the two main results for the game (N, C_{fix}) analysis in Theorem 1 and 2. We start with the most profitable coalition structure of the game (N, C_{fix}) .

Theorem 1. The service pooling game (N, C_{fix}) is a strictly subadditive game, and the grand coalition is the most profitable coalition structure.

Proof. In the total cost function $C_{fix}(S) = H_S + D_S = h_S \mu_S + d_S L_{q,S}$, $H_S = h_S \mu_S$ is additive, and $D_S = d_S L_{q,S}$ is strictly subadditive based on proposition 1. It is clear that C_{fix} is strictly subadditive, so (N, C_{fix}) is a strictly subadditive game, and any splitting of the grand coalition means an additional congestion cost for the entire set N . Therefore, the grand coalition N is the most profitable coalition structure for the game (N, C_{fix}) . \square

Subadditivity is a necessary condition required for the formation of the grand coalition. Theorem 1 states that there is always a benefit if two service providers or pooling systems share their service capacities as a "super-server" with a fixed service capacity. When each provider maintains its individual server with its own capacity, it has to pay its capacity cost and the individual congestion cost. In the context of resource pooling, each participant could save money in the collective congestion cost. Although the profitability of the grand coalition is affirmed by Theorem 1, the existence of stable cost allocation has not been confirmed yet. In order to encourage every provider joining the grand coalition, our interest now is to find a stable allocation rule to share the reduced congestion cost in the coalition derived from service capacity pooling. One of the important properties for the stable cost allocation analysis is the concavity of cost games in the following definition.

Definition 1. A TU-game (N, v) is concave if:

- (i) For any pair of subsets $\emptyset \subseteq S \subset T \subset N$, and any agent $l \in N \setminus T$, $v(S \cup \{l\}) - v(S) \geq v(T \cup \{l\}) - v(T)$; or
- (ii) For any pair of subsets $\emptyset \subseteq S, T \subset N$, $v(S \cup T) + v(S \cap T) \geq v(S) + v(T)$; or
- (iii) For any pair of subsets $S \subset T \subset N \setminus R$, $v(S \cup R) - v(S) \geq v(T \cup R) - v(T)$.

If (N, C_{fix}) is concave, the non-emptiness of the core and the existence of population monotonic allocation scheme (PMAS, which presents a dynamic solution for all sub-games) would be ensured. Unfortunately, the following example illustrates that the game (N, C_{fix}) is not concave in general.

Example. Consider the case where $h = 0$ (because of the additivity of capacity holding cost, this assumption does not affect the final result), $d = 1$ and $f = 0.2$, a set $N = \{1, 2, 3\}$ of three service providers, and the other parameters as defined in Table I.

We use the same arrival and service rates as that in the example of [10]. There are big differences among service utilizations of each agent in this situation. By choosing $S = \{1\}$, $T = \{1, 3\}$ with $l = 2$, we obtain $C_{fix}(S \cup \{l\}) - C_{fix}(S) =$

TABLE I. 3 PLAYERS POOLING GAME

| Agents | Parameters | | |
|--------|-------------|---------|-------|
| | λ_i | μ_i | C_i |
| 1 | 9 | 10 | 4.86 |
| 2 | 5 | 10 | 0.3 |
| 3 | 2 | 10 | 0.03 |

TABLE II. COALITIONAL AND DISTRIBUTED COSTS

| S | $C_{fix}(S)$ | $\sum_{i \in S} x_i^{sh}$ |
|---------------|--------------|---------------------------|
| $\{1, 2, 3\}$ | 0.37 | 0.37 |
| $\{1, 2\}$ | 0.98 | 1.33 |
| $\{2, 3\}$ | 0.11 | -1.52 |
| $\{1, 3\}$ | 0.4 | 0.91 |

$-3.88 < C_{fix}(T \cup \{l\}) - C_{fix}(T) = -0.03$, meaning that the game (N, C_{fix}) is not concave in this setting.

In Table II, the total cost of each coalition and the corresponding costs of Shapley-value allocation, which has $x_1^{sh} = 1.88$, $x_2^{sh} = -0.54$ and $x_3^{sh} = -0.97$, are computed for every coalition $S \in N$ with $|S| \geq 2$. We find that $C_{fix}(\{1, 2\}) = 0.37 < x_1^{sh} + x_2^{sh} = 1.88 + (-0.54) = 1.34$, also, $C_{fix}(\{1, 3\}) = 0.4 < x_1^{sh} + x_3^{sh} = 1.88 + (-0.97) = 0.91$. It means that the Shapley-value allocation is not stable in this case, although the pooling contribution of each agent is well shown by this allocation. Meanwhile, the general proportional allocation rule $x_i^p = p_i C_{fix} / \sum_{j \in N} p_j$, depending on the initial individual service capacity ($p_i = \mu_i$) or the own customer arrival rate ($p_i = \lambda_i$), also could not guarantee the stability of the grand coalition N .

In order to keep all agents staying in the coalition, we should find at least one stable cost allocation if it exists. Unfortunately, we could not give an explicit stable cost-sharing solution for the game (N, C_{fix}) . However, we prove the existence of the stable cost allocation solutions. To prove the non-emptiness of the core, we use the "Bondareva-Shapley Theorem" [17] (B-S Theorem), which is well known in non-empty core demonstration for TU-games: "A TU-game (N, v) has a non-empty core if and only if it is balanced". The following definitions are the relevant notions of balancedness.

Definition 2. A collection B on N is a *balanced collection*, if for any $S \in B$, it exist its *weights* β_S to make $\sum_{S \in B} \beta_S \mathbf{1}_S = \mathbf{1}_N$. This equation is equivalent to $\sum_{S \ni i} \beta_S = 1$, for any $i \in N$.

Definition 3. A cost TU-game (N, v) is a *balanced game*, if for any balanced collection B on N , we have $v(N) \leq \sum_{S \in B} \beta_S v(S)$ ($v(N) \geq \sum_{S \in B} \beta_S v(S)$ for profit games).

For the balancedness proof of the game (N, C_{fix}) , we use the following proposition.

Proposition 2. The expected waiting time $W_q(\lambda_S, \mu_S)$ in the queue is a decreasing and convex function in μ_S with a fixed λ_S for $\mu_S > \lambda_S$.

Proof. We have $\partial W_q / \partial \mu_S = -\lambda_S(1 + f)(2\mu_S - \lambda_S) / [2\mu_S^2(\mu_S - \lambda_S)^2] < 0$ and $\partial^2 W_q / \partial \mu_S^2 = \lambda_S(1 + f)[\lambda_S^2 + 3\mu_S(\mu_S - \lambda_S)] / [\lambda_S^3(\lambda_S - \lambda_S)^3] > 0$, for $\mu_S > \lambda_S$. So, $W_q(\mu_S)$ is decreasing and convex in μ_S . \square

With Proposition 2, we are prepared to prove the non-emptiness of the core for the game (N, C_{fix}) .

Theorem 2. The service pooling game (N, C_{fix}) has a non-empty core, and there are infinitely many solutions in the core if $n > 1$.

Proof. The game (N, C_{fix}) could be divided into two games: the game (N, H_{fix}) with $H_{fix}(S) = h\mu_S$, which is a linear game, and the game (N, D_{fix}) with $D_{fix}(S) = dL_q(\lambda_S, \mu_S) = d\lambda_S W_q(\lambda_S, \mu_S)$. It is clear that (N, H_{fix}) has a core with one core allocation. Now, we will prove that the game (N, D_{fix}) has a non-empty core by B-S Theorem.

For any balanced collection B on N , we have:

$$D_{fix}(N) = d\lambda_N W_q(\lambda_N, \mu_N) \\ = d\lambda_N W_q(\lambda_N, \sum_{S \in B} \beta_S \mu_S \frac{\lambda_N}{\lambda_S} \cdot \frac{\lambda_S}{\lambda_N}) \quad (5a)$$

$$\leq d\lambda_N \cdot \sum_{S \in B} \beta_S \frac{\lambda_S}{\lambda_N} W_q(\lambda_N, \mu_S \frac{\lambda_N}{\lambda_S}) \quad (5b)$$

$$= \sum_{S \in B} \beta_S d\lambda_S W_q(\lambda_N, \mu_S \frac{\lambda_N}{\lambda_S}) \\ = \sum_{S \in B} \beta_S d\lambda_S \left[\frac{1+f}{2} \lambda_N^{-1} \frac{1}{\rho_S^{-1}(\rho_S^{-1} - 1)} \right] \\ < \sum_{S \in B} \beta_S d\lambda_S W_q(\lambda_S, \mu_S) \quad (5c) \\ = \sum_{S \in B} \beta_S D_{fix}(S).$$

From the definition of a balanced collection, there is $\sum_{S \in B} \beta_S \mu_S = \mu_N$ to guarantee the equality in (5a). Meanwhile, the inequality in (5b) holds by the convex property of $W_q(\lambda_S, \mu_S)$ in Proposition 2 and $\sum_{S \in B} \beta_S \lambda_S = \lambda_N$ based on Definition 2. Since S is a subset of N , the inequality in (5c) holds.

Thus, the game (N, D_{fix}) is a balanced game. According to the B-S Theorem, the game (N, D_{fix}) has a non-empty core. Using Lemma A.2 of [12], we could simply state that: if $n > 1$, the game (N, C_{fix}) has infinitely many core allocations. The proof of the theorem is completed. \square

From Theorem 2, we could conclude that the service pooling game (N, C_{fix}) with a fixed service capacity always has a cost allocation to guarantee the stability of the grand coalition. Thus, the grand coalition is the most profitable coalition solution in the context of capacity pooling. For the further results of the individual pooling costs, the explicit numerical solution of the game (N, C_{fix}) could be computed by a Linear Programming method, e.g., the Equal Profit Method in [14] or the Nucleolus computing [18].

IV. COST-SHARING IN M/G/1 POOLING SYSTEM

Obtaining a stable allocation is one important objective of the service pooling problem. We denote the cost allocations of the grand coalition by $x = \{x_i, \dots, x_n\}$ for the cost TU-game (N, C_{fix}) . The stable allocation rule should be efficient $\sum_{i \in N} x_i = C_N$ and justified by coalition rationality $\sum_{i \in S} x_i \leq C_S$ for any subset S in N . In this section, we

propose a reasonable expression for Equal Profit Method, and compare the expression proposed with another rules and its original form by numerical experiments.

A. Cost allocation rules

1) *Shapley-value*: The *Shapley-value*, introduced by Shapley in 1952 [19], is a popular allocation concept in cooperative game theory. It provides us with a well-known fair allocation rule for cooperative games, but there is no general property to keep the stability of the grand coalition. The Shapley-value is defined as the average marginal cost of each cooperative subset for each participant, and it is given by

$$x_i^{sh} = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [C_{S \cup \{i\}} - C_S], \quad (6)$$

where $C_{S \cup \{i\}} - C_S$ is the marginal cost of provider i as the last player joining the coalition S . Shapley-value is the unique allocation rule satisfying the four desirable properties: anonymity, efficiency, additivity and dummy player property. In order to verify the Shapley-value is a core allocation for a cost cooperative game (N, v) , it is sufficient to test the concavity of the characteristic function v , which has been proved by [20]. Unfortunately, the game (N, C_{fix}) defined here is non-concave and the Shapley-value couldn't guarantee the stability of grand coalition in general.

2) *Tau-value*: The set of cost allocations, defined by the lower bound $M_i(N, C_{fix}) = C_N - C_{N \setminus \{i\}}$ and the upper bound $m_i(C_{fix}) = \max_{S: i \in S} \{C_S - \sum_{j \in S \setminus \{i\}} M_j(N, C_{fix})\}$ of the shared cost, includes all the core allocations. The τ -value, defined by Tijs in 1981 [21], is a cost allocation defined as

$$x_i^\tau = \alpha m_i(C_{fix}) + (1 - \alpha) M_i(N, C_{fix}), \quad (7)$$

with the unique $\alpha \in [0, 1]$ calculated by the efficiency of x^τ . It is a special linear combination of $M_i(N, C_{fix})$ and $m_i(C_{fix})$. For two player games, the τ -value is equal to the Shapley-value and it presents a stable cost allocation for all quasi-balanced two player games. Although we could not give an explicit demonstration of its stability, the τ -value proposes stable results in the following experiments.

3) *Nucleolus*: The *excess* of a cost allocation x for a coalition $S \subseteq N$ is defined as

$$e_{x,S} = C_S - \sum_{i \in S} x_i, \quad (8)$$

which is been used to measure *unhappiness* of the coalitions by the *lexicographic ordering* comparison for all possible allocations. When the coalition cost is allocated by x , S is more satisfied with a higher $e_{x,S}$. Based on the concept of *minimized maximum unhappiness*, the *Nucleolus* was introduced by Schmeidler in [18]. Although Schmeidler didn't define it explicitly, its uniqueness has been proved by Driessen in 1969 [22]. It is a stable allocation rule for all TU-games with a non-empty core, and coincides with dummy player property, zero independent and reduced-game property.

4) *Equal Profit Method*: Another interesting definition of "fair" is the *equal profit* concept, which defined by Frisk et al. in 2010 [14]. They describe the cost TU-games as a linear programming problem to minimize the gap of the relative savings $r_i = (C_i - x_i)/C_i$.

$$\begin{aligned} \min f(x) \\ \text{s.t } f(x) &\geq \max\{r_i - r_j\}, \forall i, j \in N \\ \sum_{i \in S} x_i &\leq C_S, \forall S \subset N \\ \sum_{i \in N} x_i &= C_N, \end{aligned} \quad (9)$$

where the results x^{EPM} defined as the core allocation calculated by *Equal Profit Method* (EPM). It seems to be a "fair" and stable allocation rule considering the most closed relative savings for each individual, despite the different contribution of each participant to the grand coalition N . This "fair" definition just considers the individual factor. In some cases, the participant, which has a low individual payment, might do not want to join the coalition in order to protect its individual information, although it could get a same level relative saving with others.

5) *EPM based on Contribution Weights*: Now, we define the *relative contribution weights* $w_i = \lambda_i/(\mu_i - \lambda_i)$ for each participant in our service pooling game. The customer arrival rate λ_i describes the individual requirement of each participant, and the system idle capacity $\mu_i - \lambda_i$ presents the individual contribution to the collaboration. Then, we propose the new relative saving formula as

$$rw_i = w_i * \frac{C_i - x_i}{C_i} = \frac{\lambda(C_i - x_i)}{(\mu - \lambda)C_i}. \quad (10)$$

Thus, we get the *EPM with Contribution Weights* (EPMCW) as

$$\begin{aligned} \min f(x) \\ \text{s.t } f(x) &\geq \max\{rw_i - rw_j\}, \forall i, j \in N \\ \sum_{i \in S} x_i &\leq C_S, \forall S \subset N \\ \sum_{i \in N} x_i &= C_N. \end{aligned} \quad (11)$$

With the constraints defined in equation (11), EPMCW presents a stable allocation rule by solving the linear programming problem above, if the core is not empty. The contribution of each provider has been considered in the *relative contribution weights* w_i , and the results are easily controllable by w_i with different definitions, e.g., $w_i^* = \lambda_i^2/(\mu_i - \lambda_i)$.

B. Numerical results and analysis

We illustrate the previous concepts in three typical service pooling cases of 6 service companies, which have equal individual service capacities. To do so, we consider the three following sets of data with different $h \in [0, 1]$ (the variation of the three system parameters h, d, f have similar impacts on the results). The first case presents a set of companies

TABLE III. CUSTOMER ARRIVAL RATES AND SYSTEM PARAMETERS

| Initial data | | | | | Customer arrival rates | | | | | |
|--------------|--------|-----|-----|-----|------------------------|---|---|-----|-----|-----|
| μ | h | d | f | No. | 1 | 2 | 3 | 4 | 5 | 6 |
| 10 | [0, 1] | 2 | 0.2 | 1 | 2 | 3 | 3 | 4 | 3.5 | 4 |
| - | - | - | - | 2 | 7 | 9 | 8 | 7 | 8.5 | 9.5 |
| - | - | - | - | 3 | 2 | 7 | 4 | 7.5 | 9 | 3 |

with low server utilizations. It means that all the companies in this set are not very efficient. There are both the low efficient companies and the busy companies in the second case. And the third one only consists of the busy companies. The initial parameters for each set are listed in Table III.

Fig. 2, 3 and 4 reveal the distributed costs, using different preceding concepts described in IV-A, for our game (N, C_{fix}) in the cases above under $h = 0$, i.e., the quality driven case. We select two different relative contribution rates for EPMCW calculation: w_i for EPMCW1 and w_i^* for EPMCW2. From these figures, we find that the results given by the τ -value

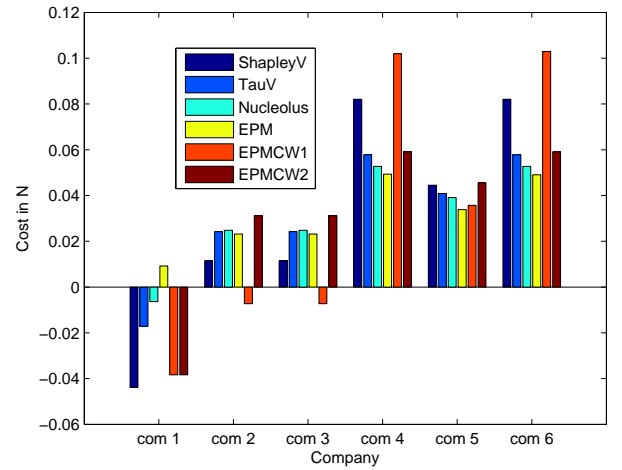


Fig. 2. Cost allocations of case 1 under $h = 0$ by Shapley-value, Tau-value, Nucleolus, EPM, EPMCW1 and EPMCW2

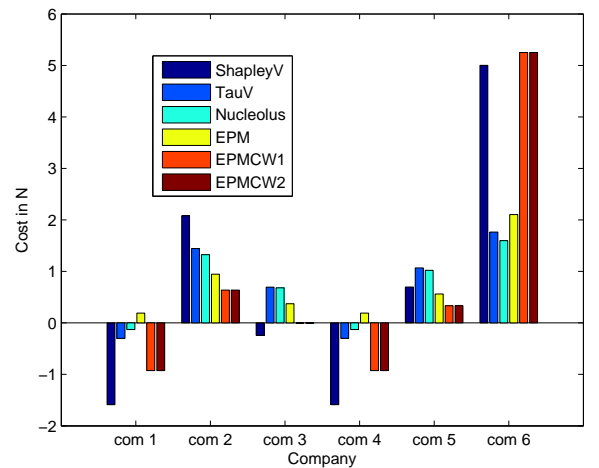


Fig. 3. Cost allocations of case 2 under $h = 0$ by Shapley-value, Tau-value, Nucleolus, EPM, EPMCW1 and EPMCW2

and the nucleolus are very close, especially in the third case. The Shapley-value reflects the contributions of each provider for all the possible coalitions. Unfortunately, the corresponding allocations are not stable in all the three cases, e.g., $x_1^{sh} + x_5^{sh} = 1.2811 \geq C_{\{1,5\}} = 0.8067$ in the case 3.

EPM provides a stable allocation rule, which is a little far from the others for several companies, particularly for the company with relative large or small contribution to the coalitions. The goal or the "fair" defined by EPM is to minimize the gap between the relative earnings of players, and the constraints consists of all stable requirements. Thus, the companies with a special contribution might not satisfy the similar relative saving. For example, the first company in Fig. 4, which has a low individual operating cost, will pay nothing by EPM computing, but it could earn money in other allocations.

TABLE IV. RELATIVE SAVINGS r_i FOR CASE 1 WITH DIFFERENT RULES UNDER $h = 0$ AND 0.5

| $h = 0$ | Shapley | Tau-v | Nucl. | EPM | E-CW1 | E-CW2 |
|-----------|---------|--------|--------|-------|--------|--------|
| Com 1 | 173.0% | 128.5% | 110.5% | 84.7% | 164.0% | 164.0% |
| Com 2 | 92.6% | 84.3% | 84.9% | 85.0% | 104.7% | 79.8% |
| Com 3 | 92.6% | 84.3% | 84.9% | 85.0% | 104.7% | 79.8% |
| Com 4 | 74.3% | 81.9% | 83.5% | 84.6% | 68.1% | 81.5% |
| Com 5 | 80.3% | 81.9% | 82.7% | 85.1% | 84.2% | 79.9% |
| Com 6 | 74.3% | 81.9% | 83.5% | 84.6% | 68.1% | 81.5% |
| Stable | No | Yes | Yes | Yes | Yes | Yes |
| $h = 0.5$ | Shapley | Tau-v | Nucl. | EPM | E-CW1 | E-CW2 |
| Com 1 | 2.05% | 1.52% | 1.31% | 1.94% | 1.94% | 1.94% |
| Com 2 | 2.77% | 2.52% | 2.51% | 3.14% | 3.14% | 3.14% |
| Com 3 | 2.77% | 2.52% | 2.51% | 3.14% | 3.14% | 3.14% |
| Com 4 | 4.47% | 4.93% | 5.02% | 3.94% | 3.82% | 3.82% |
| Com 5 | 3.48% | 3.55% | 3.58% | 3.94% | 4.19% | 4.19% |
| Com 6 | 4.47% | 4.93% | 5.02% | 3.94% | 3.82% | 3.82% |
| Stable | No | Yes | Yes | Yes | Yes | Yes |

EPMCW1 and EPMCW2 give closed results in Fig. 3 and 4, but they propose very different results in Fig. 1. In these figures, we find that the EPMCW2 is more closed to the Shapley-value. It seems like w_i^* presents a more suitable

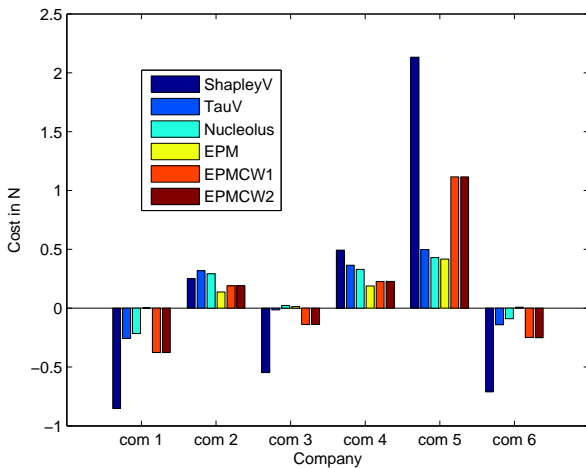


Fig. 4. Cost allocations of case 3 under $h = 0$ by Shapley-value, Tau-value, Nucleolus, EPM, EPMCW1 and EPMCW2

contribution weights than w_i defined in equation (10). With same individual service capacities used for all companies, this phenomenon could be explicated by the characteristic function C_{fix} of this service pooling game defined in equation (4). It is easy to find that the second contribution weights w_i^* are more closed to the individual operating cost C_i .

Table IV shows the relative savings r_i for each distributed cost obtained by the previous rules and compared to the individual costs in the set of low efficient companies (the case 1 in Fig. 2). When $h = 0$, the service capacity pooling brings a large relative saving to all companies. The relative difference is large between the result of EPMCW1 and EPMCW2. While the relative saving declines, $h = 0.5$, the difference between EPMCW1 and EPMCW2 reduces. If the relative saving is small enough, e.g., $h = 1$ with $r_N \leq 3\%$, the two allocations given by EPMCW and the allocation of EPM are very similar.

V. CONCLUSION

Using cooperative game theory, we have approached the cost sharing problem among a set of independent service providers in the complete service capacity pooling system.

We extended the service pooling game in M/G/1 service systems. When the service capacities are fixed, the service pooling game would be a sum of two games: an additive service holding cost game and a subadditive service congestion game. We have proved that the stable cost allocation always exists. Thus, the stable numerical cost allocation could be derived using linear programming computing for any cases. Considering the contribution and requirement of each service providers, we propose an improved expression of Equal Profit Method, the Equal Profit Method based on Contribution Weights. We have applied it in our M/G/1 service pooling cases. With numerical experiments, we found that it provides a reasonable stable allocation rule, which could reflect the contributions of each participant in the coalition.

There are possible generalizations and complementarities that can be taken in future research. In our setting, (h, d, f) are system parameters. Because of the problems based on the localization, it is also reasonable to suppose that d_i is specialized for each service provider. The expected waiting time could increase by service capacity pooling for several service providers. For this, we could not guarantee the subadditivity of the congestion cost. In this way, the grand coalition may not be the most profitable coalition structure for this new game. Thus, the coalition formation problem should be considered in the service pooling game study.

It would be also interesting to extend the analysis of this paper to the case of service systems with customers abandonment or callback. Indeed, the abandonment is a common phenomenon in service systems. The customers will abandon with the unsatisfying of the real waiting times. Thus, the stability requirement $\lambda < \mu$ is not necessary. Considering the callback of the fail/unsatisfying demand, it is necessary to propose a fair indicator to evaluate the quality of services. There are a few options here: the customers waiting times, the service times, the probability of failure, etc. It will be an interesting future research to analyze the properties of the relevant games.

REFERENCES

- [1] S. United States Trade Representative and Investment, "Service," <https://ustr.gov/issue-areas/services-investment/services>, 2014.
- [2] A. Katta and J. Sethuraman, "Cooperation in queues," CORC technical reports, TR-2005-03, Columbia University, New York, Tech. Rep., 2006.
- [3] Ç. Kayi and E. Ramaekers, "Characterizations of pareto-efficient, fair, and strategy-proof allocation rules in queueing problems," *Games and Economic Behavior*, vol. 68, no. 1, pp. 220–232, 2010.
- [4] P. Guo, M. Leng, and Y. Wang, "A fair staff allocation rule for the capacity pooling of multiple call centers," *Operations Research Letters*, vol. 41, no. 5, pp. 490–493, 2013.
- [5] E. Tekin, W. Hopp, and M. Van Oyen, "Pooling strategies for call center agent cross-training," *IIE Transactions*, vol. 41, no. 6, pp. 546–561, 2014.
- [6] O. Aksin, F. de Véricourt, and F. Karaesmen, "Call center outsourcing contract analysis and choice," *Management Science*, vol. 54, no. 2, pp. 354–368, 2008.
- [7] S. Anily and M. Haviv, "Subadditive and homogeneous of degree one games are totally balanced," *Operation Research*, vol. 62, no. 4, pp. 788–793, 2014.
- [8] P. González and C. Herrero, "Optimal sharing of surgical costs in the presence of queues," *Mathematical Methods of Operations Research*, vol. 59, no. 3, pp. 435–446, 2004.
- [9] M. Garcia-Sanz, F. Fernández, M. Fiestras-Janeiro, I. García-Jurado, and J. Puerto, "Cooperation in markovian queueing models," *European Journal of Operational Research*, vol. 188, no. 2, pp. 485–495, 2008.
- [10] Y. Yu, S. Benjaafar, and Y. Gerchak, "Capacity sharing and cost allocation among independent firms with congestion," *Production and Operations Management*, doi: 10.1111/poms.12322, 2015.
- [11] S. Anily and M. Haviv, "Cooperation in service systems," vol. 58, no. 3, pp. 660–673, 2010.
- [12] F. Karsten, M. Slikker, and G. van Houtum, "Resource pooling and cost allocation among independent service providers," *Operation Research*, vol. 63, no. 2, pp. 476–488, 2015.
- [13] —, "Analysis of resource pooling games via a new extension of the erlang loss function," BETA Working Paper 344, Eindhoven University of Technology, Tech. Rep., 2011.
- [14] M. Frisk, M. Göthe-Lundgren, K. Jörnsten, and M. Rönnqvist, "Cost allocation in collaborative forest transportation," *European Journal of Operational Research*, vol. 205, no. 2, pp. 448–458, 2010.
- [15] M. Guajardo and M. Rönnqvist, "Cost allocation in inventory pools of spare parts with service-differentiated demand classes," *International Journal of Production Research*, vol. 53, no. 1, pp. 220–237, 2015.
- [16] F. Pollaczek, "Über eine aufgabe der wahrscheinlichkeitstheorie," *Mathematische Zeitschrift*, vol. 32, no. 1, pp. 64–100, 1930.
- [17] O. Bondareva, "Some applications of linear programming methods to the theory of cooperative games," *Problemy Kibernetiki*, vol. 10, pp. 119–139, 1963.
- [18] D. Schmeidler, "The nucleolus of a characteristic function game," *SIAM Journal on Applied Mathematics*, vol. 17, no. 6, pp. 1163–1170, 1969.
- [19] L. S. Shapley, "A value for n-person games," DTIC Document, Tech. Rep., 1952.
- [20] L. Shapley, "Cores of convex games," *International Journal of Game Theory*, vol. 1, no. 1, pp. 11–26, 1971.
- [21] S. Tijs, "Bounds for the core of a game and the t-value," *Game Theory and Mathematical Economics*, pp. 123–132, 1981.
- [22] T. S. H. Driessen, *Cooperative Games, Solutions and Applications*. Kluwer Academic Publishers, 1988.